Noname manuscript No.
(will be inserted by the editor)

Social Media and Satellites

Disaster event detection, linking and summarization

Kashif Ahmad · Konstantin Pogorelov · Michael Riegler · Nicola Conci · Pål Halvorsen

Received: date / Accepted: date

Abstract Being able to automatically link social media and satellite imagery holds large opportunities for research, with a potentially considerable impact on society. The possibility of integrating different information sources opens in fact to new scenarios where the wide coverage of satellite imaging can be used as a collector of the fine-grained details provided by the social media. Remote-sensed data and social media data can well complement each other, integrating the wide perspective provided by the satellite view with the information collected locally, being it textual, audio, or visual. Among the possible applications, natural disasters are certainly one of the most interesting scenarios, where global and local perspectives are needed at the same time.

In this paper, we present a system called JORD that is able to autonomously collect social media data (including the text analysis in local languages) about technological and environmental disasters, and link it automatically to remote-sensed data. Moreover, in order to ensure the quality of retrieved information, JORD is equipped with a hierarchical filtering mechanism relying on the temporal information and the content analysis of retrieved multimedia data.

To show the capabilities of the system, we present a large number of disaster events detected by the system, and we evaluate both the quality of the provided information about the events and the usefulness of JORD from potential users viewpoint, using crowdsourcing.

Keywords Information Retrieval · Event Detection · Natural Disaster · Social Media

1 Introduction

Wide geographical coverage and high spatial and multi-spectral resolutions are the key characteristics of satellite imagery, which make it a useful source of information and support tool in different application areas. Satellite imagery has been mostly used to explore and monitor the surface of the earth; The most famous and longest-running satellite programs is called LandSat, launched with the objective of gathering Earth resource data. NASA has recently

DISI-University of Trento
Via Sammarive 5, Italy
E-mail: kashif.ahmad@unitn.it

Fig. 1 A satellite image of wildfires in Fort McMurray, Canada. Based on the image, though the wildfires can be observed, it is almost impossible to give a clear statement about its impact on environment and society (image from planet[4]).

released the data of Landsat[1], opening a number of opportunities both for the society and research, enabling the development of systems that integrate remote sensed data in different applications. Examples include automatic detection of forest fires, sandstorms and floods [22, 30], besides being used to develop and evaluate satellite image classification systems for change detection (pre- and post-disaster), for damage assessment.

However, satellite imagery also comes with interesting challenges, and its suitability to tackle a particular application depends on different factors, such as geographical coverage, resolutions and temporal frequency. It is well known that satellite images have a low temporal frequency, which makes its use questionable in time-sensitive applications. Furthermore, satellite imagery only give a birds'-eye view of an actual event [45, 14]. For instance, as shown in Figure 1, wildfires can be detected in the satellite image of Fort McMurray, Canada, taken from Planet 4-band satellite[2], but this information does not report the impact the wildfire had on people's life.

On the other hand, in recent years social media has emerged as an important source of information and rapid communication in emergency situations [62]. As demonstrated in [53], there are a number of emergency situations in which news agencies could not provide information at all or in time, simply due to the lack of reporters spread all over the world. In such circumstances, social networks play an important role in breaking and disseminating news [16].

Inferring events, in general, through information available in social media has also been an area of interest for the researchers [44], with strong focus on the capability to detect, track and summarize the contents of the underlying events. The use of social media also comes

[1] http://landsat.usgs.gov/
[2] https://www.planet.com

with challenges, including the collection and management of data, as well as the validation of its reliability.

A rather recent trend is to gather information about critical events, such as natural disasters, by combining data available on the social networks with remote sensed data[5] [6], which confirms the interest of the multimedia research community in the topic.

In this paper, we present the JORD system, which is to the best of our knowledge the first one that collects, analyzes and combines multi-modal information (text, images and videos) about disasters from different social media platforms, and links it with remote sensed data in real-time[7]. It also provides query refinement by automatically generalizing it in all local languages that are relevant to the position of a disaster. Moreover, to ensure the quality of retrieved multimedia data, we propose multi-modal content analysis and temporal filtering. In order to obtain positioning information necessary to retrieve and link satellite imagery to the underlying events, we extract the GPS coordinates of the places and city names mentioned in the tweets relying on natural language processing (NLP) techniques. The system is also equipped with a novel methodology for identifying the areas hit by the disaster in complex satellite imagery. For evaluation purposes, we have conducted a crowdsourcing campaign with a large number of users, asking them to share their feedback about the retrieved contents and the system itself.

With a system, that combines multimedia mining [38], retrieval [13] and linking [21] methods, we are able to tell a much clearer and more useful story to the users. In summary, we can synthesize the main features of JORD as:

(i) It collects data about events autonomously and automatically in real-time from a disaster database (i.e., when JORD is running and an event occurs, it will continuously gather new information from social media to enhance the event information).
(ii) JORD is able to generate queries in local languages spoken in the area hit by the underlying disaster.
(iii) JORD automatically filters irrelevant information in a hierarchical way relying on temporal information and content analysis of the retrieved data.
(iv) JORD combines social media and satellite imagery in a novel way, and provides a more detailed event description to the users.
(v) It is equipped with a novel method for linking and retrieving satellite imagery with the events by analyzing the tweets text to identify and extract GPS coordinates of the areas struck by the disaster.
(vi) JORD also consists of a novel framework for flood detection in satellite images as a use-case of the disaster event detection in satellite imagery.

This paper is an extension of our previous work [4]. Our extended contribution is threefold: (i) we introduce a novel method for retrieving and linking satellite imagery with events by extracting the GPS coordinates of the places and cities affected by a disaster from the tweet's text; (ii) we extend content analysis to other types of data (tweets' text) (iii) a novel methodology relying on Generative Adversarial Networks (GANs) is proposed for flood detection in satellite imagery as a use-case for disaster detection in satellite images.

The rest of the paper is organized as follows: Section 2 provides a detailed description of the related work by analyzing the importance of remotely sensed data in different appli-

[5] http://www.multimediaeval.org/mediaeval2017/multimediasatellite/index.html

[6] `http://www.acmmm.org/2016/wp-content/uploads/2016/03/ACMMM16_GC_Sky_and_the_Social_Eye_latest.pdf`

[7] Real-time in the context of information retrieval means in our case that JORD continuously monitors various information sources and retrieves the information as soon as a query match is found.

cations with particular emphasis on natural disasters, and characteristics of social media. In Section 3, we present the proposed system, describing the overall architecture, the nature of the disasters retrieved by JORD from social media, along with a detailed description of the methodologies proposed for the content analysis of retrieved multimedia data, linkage with satellite imagery and the evaluation of the system through a crowdsourcing study to gather real user feedback. In Section 4, a detailed analysis and discussion of the experimental results is conducted with a focus on the flood detection use case. Section 5 concludes the paper and presents the potential improvements and future work directions.

2 Related Work

Remote Sensed Data

Since the launch of Landset 1[8], formerly known as Earth Resources Technology Satellite (ERTS), satellite imagery has been widely used to address a diversified set of applications including meteorology, fishing industry, agriculture, forestry, geology, regional planning, education and warfare [12]. However, the use of satellite imagery in an application depends on a number factors, such as spatial and spectral resolution, coverage and cloud cover.

Over the last few years, satellite data has also been used in disaster management to analyze its impacts on the environment. The wide geographical coverage and multi-spectral resolution make satellite imagery an important source of information in these scenarios. For instance, according to the authors in [27], the disaster management process can be roughly divided into 4 different phases, and satellite data is equally useful in all of them. These phases include reduction, readiness, response, and recovery. Moreover, a number of international cooperation mechanisms and organizations have been established to help and support in disaster management, which heavily rely on remote sensed data [58,27].

As of today, the literature about disaster detection through satellite data is rather limited. Amit et al. [6] propose a Convolutional Neural Networks (CNNs) based approach for the detection of disasters, such as landslides and floods, in satellite imagery. A similar approach is adopted in [29], where a deep model is trained on aerial photos captured through unmanned aerial vehicles (UAV). CNN features are also exploited by Liu et al. [36] for the representation of landslide images in their disaster recognition system. Deep features have been utilized for the detection of flooding [9] and also wildfire and earthquake [10] events in social media and satellite imagery by Bischke et al. showing the promising results in a combination of multi-modal data sources.

Thanks to the growing interest in the topic, a benchmarking challenge has been introduced at MediaEval 2017, addressing flood detection in satellite images [11]. Benjamins's et al. [11] approach flood detection in satellite imagery as a segmentation problem relying on three different variations of a deep model [51]. In details, the final convolutional layer of the model is replaced with an up-sampling layer relying on bi-linear interpolation to re-scale the down-sampled feature maps into original patch size. Subsequently, a softmax layer is used to classify the pixels into flooded and non-flooded regions. Similarly, in [31], an approach based on the concept of convolutional deep model with dilated convolution is proposed to deal with the segmentation and classification of satellite image patches into flooded and non-flooded regions. In total, four different models with different number of dilated convolutional layers are used. Avgerinakis et al. [32] use Mahalanobis distances with stratified co-variance estimates along with morphological post-processing to this aim.

[8] https://landsat.gsfc.nasa.gov/landsat-1/

It is to be noted that satellite data also have limitations. The low temporal frequency is one of the biggest hurdles, particularly in those contexts where a prompt feedback is required. However, the availability of the images before and after a disaster[9] could contribute to a more detailed description of the event, when combined with other information sources as, for example, social media.

Social Media

The huge amount of content shared through social networks provides useful information for many applications and research studies in different fields, such as economics, sociology and computer science. The attractiveness of social media is indisputable, as it is a powerful medium for the dissemination of information in a lot of domains [19], and can be regarded as an effective medium of mass communication [53], also in emergency situations.

In this regard, a common practice is to infer events from the information shared through social media. To this aim, Popescu and Pennacchiotti [48] extract a list of actors, musicians, politicians and sports men from Wikipedia to be used for crawling Twitter, and detect controversial events about them. Other approaches rely on unsupervised frameworks for the detection of social events in Twitter [7]. Mathiodakis et al. [39] use clustering techniques on bursty keywords to detect trends in Twitter. Meladianos et al. [40] propose a methodology for sub-event (i.e., key moments of an event) detection in Twitter streams using the concept of graph degeneracy. In [25], a statistical approach relying on tweets, and the frequency of links, inserted by users in their tweets, is proposed to detect social events.

A number of works in this regard also exploit Twitter data to detect and analyze emergency situations and disaster events. For instance, Li et al. [35] propose a method to detect crime and disaster events in Twitter's text streams. In [50], tweets are analyzed to detect earthquakes in Japan, where key words, such as earthquake and typhoons, are used to crawl Twitter. In [8], a method for the detection of earthquakes in tweets is proposed, where a graph-based clustering technique has been utilized to target geo-located communities in Twitter. In [17], Twitter is used as a social sensor to capture information about a natural disaster from users in real time. In [20], a concept derived from seismology, originally developed to detect seismic phases, is used for earthquake detection in Twitter text streams. The authors monitor a rapid increase in the tweets containing words *earthquake* relying on a short-term-average over long-term-average (STA/LTA) algorithm. More recently, Xu et al. [59] proposed a participatory sensing-based model for collecting information about disaster events in micro-blogs. In [55], the authors examine the use of social media, Twitter in particular, in emergency situations considering a number of factors, such as time and location of the use, and type of user (e.g., general public, journalist and add agencies etc.,).

Besides Twitter, other works have also tried to exploit other social media platforms to detect such events, as for example Flickr [57,1]. In this regard, most of the existing works target social events and daily life activities [57,2].

More recently, a benchmark initiative has been established to detect flood-related images in social media [11]. In the response to the task, a number of interesting solutions have been proposed for the classification of flooded and non-flooded images [31,32,42,3]. In [5], the classification results of different classifiers trained on different CNN models are combined in two late fusion methods. Moreover, user tags, geo-location information and

[9] `http://www.satimagingcorp.com/applications/environmental-impact-studies/natural-disasters/`

descriptions of an image are also used, as an additional information to support the visual features. Benjamin et al. [11] also rely on an image representation scheme deriving benefits from deep architectures. They extract features from two deep architectures, DeepSentiBank [15] and X-ResNet [28], and a Support Vector Machine (SVM) is used for classification purposes. Other works rely on hand-crafted visual features [41,63]. For instance, in [43], a combination of CEDD, CL and GABOR features are used along with meta-data.

3 Proposed System

As shown in Figure 2, JORD consists of four main components (highlighted in different colors with corresponding labels): (i) query refinement, (ii) multimedia data retrieval from social media, (iii) temporal and content-based filtering of the retrieved multimedia content, and (iv) linking social media data with remote-sensed data. In the query refinement phase, we generate new queries in local languages spoken in the areas struck by the disaster. Subsequently, we crawl different social media platforms to collect as much information as possible. The data retrieval phase is followed by a filtering stage, where we analyze and process the retrieved content. Next, we extract the geo-location information from images and tweets, which are then used to retrieve the satellite images. Finally, the satellite images are analyzed and processed to detect the underlying disaster event. In the next subsections, we provide a detailed descriptions of these phases.

Sensing New Events (Natural Disasters)

As aforesaid, one of the main advantages of JORD is the capability to collect information about natural and technological disasters in real time, which basically means that if JORD is once started, it will continue collecting and linking events as long as they occur. To this aim, the proposed system extracts a list of natural and technological disaster events from the EM-DAT database [24] in real-time. This means that as soon as a new event occurs in the database, JORD starts collecting and linking information about it. EM-DAT is an international disaster database (supported by the World Health Organization - WHO) that provides information of natural and technological disasters that have occurred all over the world. Table 1 provides a list of some samples events sensed and analyzed by our system. It is to be noted that JORD is able to collect, link and analyze an unlimited number of events depending on the processing and storage resources, and can operate live as an quasi autonomous system, and on demand, namely controlled by a user.

Query Refinement and Translation

It is observed that, during a natural disaster, the local community usually initiates the process of spreading the news about the event by tweeting, posting, commenting and sharing information through the respective social media platforms. Furthermore, the local community tends to share information in their local languages. Based on these observations, we believe that it would be advantageous to crawl social media with queries in local languages. Thus, our system automatically generates queries in all local languages that are relevant to the position of a disaster. For query generation in local languages, JORD relies on the information provided in the EM-DAT database. For each disaster entry in the database, EM-DAT

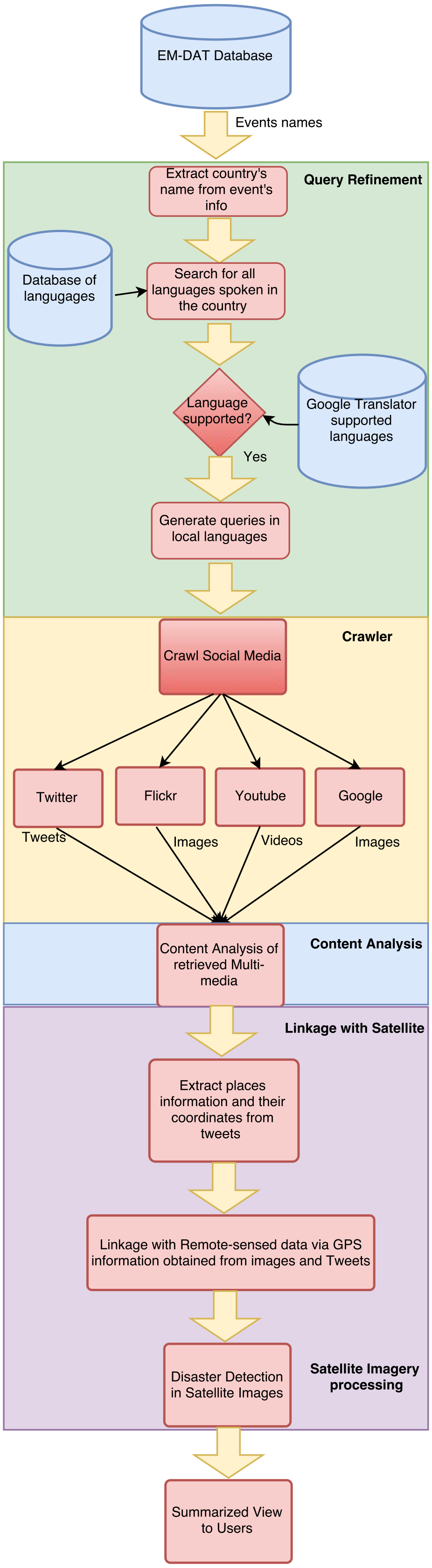


Fig. 2 Block diagram of the proposed system. Overall, the system is composed of four different phases each highlighted in different color.

Table 1 A list of examples for natural and technological disasters retrieved by JORD.

Event	Location	Time Period	Event	Location	Time Period
Earthquake	Italy	August 2016	Floods	Laos	August 2016
Earthquakes	Esmeraldas, Ecuador	May 2016	Landslides	Kegalle district, Sri Lanka	May 2016
Cyclone Roanu	Bangladesh	May 2016	Landslides	West regions of Uganda	May 2016
Tornadoes	Oklahoma, United States	May 2016	Floods	Kilinochchi district, Sri Lanka	May 2016
Thunderstorms	Bangladesh	May 2016	Landslide	Sibolangit, Indonesia	May 2016
Landslide	Rwanda	May 2016	Floods	Ethiopia	April 2016
Landslide	Uganda	May 2016	Mudslide	Taining district	May 2016
Severe weather	Haiti	May 2016	Wildfires	Alberta province, Canada	May 2016
Thunderstorms	Uruguay	April 2016	Flash flooding	Texas, United States	April 2016
Floods	Port-au-Prince, Haiti	April 2016	Floods	Southern China	April 2016
Thunderstorms	Myanmar	April 2016	Floods	Assam, Nagaland, India	April 2016
Thunderstorms	China	April 2016	Drought	India	April 2016
Drought	Timor-Leste	April 2016	Floods	Saudi Arabia	April 2016
Earthquake	Kumamoto, Japan	April 2016	Storm	Dolores, Uruguay	April 2016
Earthquake	Ecuador	April 2016	Floods	Santiago region, Chili	April 2016
Flash floods	Yemen	April 2016	Earthquake	Kumamoto, Japan	April 2016
Earthquake	Pakistan	April 2016	Floods	Ethiopia	April 2016
Storm Katie	France and UK	March 2016	Floods	KpK Pakistan	April 2016
Severe weather	United States	March 2016	Drought	India	March 2016
Floods	Kashmir, Pakistan	March 2016	Severe weather	United States	March 2016
Floods	Indonesia	March 2016	Floods	China	March 2016
Coal mine explosion	Lougansk, Ukraine	May 2016	Shipwreck	Libya	April 2016
Thunderstorms	Uruguay	April 2016	flooding	Texas, United States	April 2016
Floods	Haiti	April 2016	Floods	Southern China	April 2016
Thunderstorms	Myanmar	April 2016	Floods	Assam, India	April 2016
Drought	Timor-Leste	April 2016	Explosion in a plant	Mexico	April 2016
Shipwreck	Lybia	April 2016	Shipwreck	Mynamar	April 2016
Plane crash	Papua New Guinea	April 2016	Storm Katie	France and UK	March 2016
Floods and landslides	Pakistan	March 2016	Earthquake	Tainan, Taiwan	Feb. 2016
Earthquake	Spain and Morocco	Jan. 2016	Floods	China	Jan. 2016
Snowstorm	East coast, United States	Jan. 2016	Earthquake	Qinghai province, China	Jan. 2016
Wildfires	Spain	Dec. 2015	Tornadoes	South of United States	Dec. 2015
Floods	Kenya	Dec. 2015	Cyclone Chapala	Yemen	Nov. 2015
Plane crash	South Sudan	Nov. 2015	Floods	Somalia	Oct. 2015
Floods	Nigeria	Sept. 2015	Wildfires	California, United States	Sept. 2015
Floods	Ibaraki (Japan)	Sept. 2015	Landslides	Kaski, Nepal	July 2015
Earthquake	Pakistan	Oct. 2005	Cyclone Winston	Fiji	Feb. 2016
Wildfires	Greece	July 2015	Floods	Myanmar	July 2015

Table 2 Sample queries generated by our system in local languages.

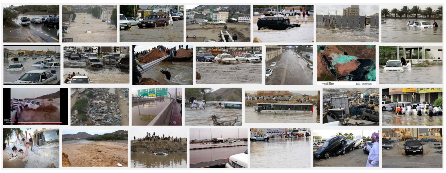
Original Query	System-generated Query	Translated to
Floods Saudi Arabia	الفيضانات المملكة العربية السعودية	Arabic
Earthquake Kumamoto (Kyushu Isl.) Japan	地震熊本（九州ISL。）日本	Japanese
Storm and flood Dolores Uruguay	Tormentas e inundaciones Dolores Uruguay	Spanish
Earthquake Ecuador	terremoto de Ecuador	Spanish
Shipwreck with migrants Libya	غرق سفينة مهاجرين مع ليبيا	Arabic
Cyclone Chapala Yemen	الإعصار تشابالا اليمن	Arabic
Floods Ibaraki Japan	洪水茨城日本	Japanese
Droughts in Tharparkar Pakistan	تھرپارکر پاکستان میں خشک سالی	Urdu
Flash floods Yemen	السيول اليمن	Arabic
Earthquake in central Italy	Terremoto in Italia centrale	Italian
Earthquake in Khyberpukhtoonkhwa Pakistan	کہبرپختونخوا پاکستان میں زلزلہ	Urdu

contains the time and location information. To this aim, the Google Translator Api[10] and a database of spoken local languages per country are integrated in the framework. JORD automatically extracts country names and the languages spoken in that country from the EM-DAT and the local database, respectively. Subsequently, Google Translator is used to translate the original queries in local languages spoken in the area. Table 2 provides some sample queries generated by our system in different languages spoken in different parts of the world.

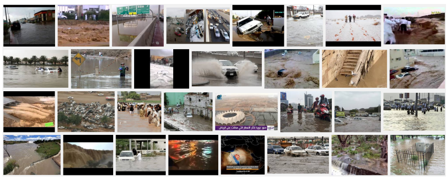
The translated queries are used in the next processing block of JORD, which is responsible for collecting multimedia data. We observed that using the translated and the original

[10] https://cloud.google.com/translate/docs/

queries results in a larger amount of retrieved data per query (for some events, a search based on only English queries results in very little or none results). Moreover, the content retrieved with translated queries are noticed to be more relevant and accurate compared to queries in English. As an example, Figure 3 shows some sample images retrieved with original and translated queries. It can be seen that the list belonging to the original query has some irrelevant images, while the images retrieved with the translated query are mostly relevant to the underlying event. Similarly, in the case of tweets, queries in local languages retrieve more accurate information. As another example, in Table 3, top 5 tweets retrieved by our system, for recent floods in Saudi Arabia, with both original (information taken from EM-DAT database) and translated queries (Arabic) along with meanings of tweets have been provided. As can be seen, the tweets retrieved with original queries are mostly irrelevant (e.g., most of them are reporting about Saudi Arabia's aid to flood victims in different parts of the world). On the other hand, the tweets retrieved with translated queries, which are expected to be posted by the local community in the local language, provide more accurate and relevant information about the event.



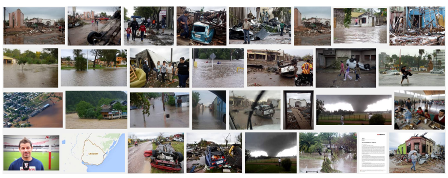
(a) Images for the event Floods in Saudi Arabia retrieved with the English query.



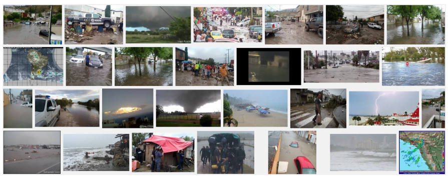
(b) Results for Floods in Saudi Arabia retrieved with the translated query.

(c) Earthquake Kumamoto (Kyushu Isl.) Japan retrieved using the English query. Some of the images seem not relevant (a bear).

(d) Image for Earthquake Kumamoto (Kyushu Isl.) Japan retrieved with the translated query. All images contain relevant information about the earth quake.



(e) English query results for Storm and flood Dolores Uruguay. Some images seem not relevant (football player).



(f) Storm and flood Dolores Uruguay images retrieved with the translated queries. All images are related to the event.

Fig. 3 Sample images retrieved with original queries in English and the translated ones. For two out of three events the English query results contain non-relevant images. This is an indicator that the translated queries can significantly improve the quality of the results.

Table 3 Top 5 Tweets retrieved with original English query and queries generate by JORD in Arabic.

Tweets with English query	Tweets with translated query	Meaning of Arabic tweets
Saudi Arabia provides 200 tons of relief and aid to those affected by the floods in Sudan	أمطار غزيرة تغسل جدة ا لمملكة العربية السعودية، شار	Heavy rain washed Jeddah, Saudi Arabia, street flooding spread
This is my pic of the day: Camels stuck in SaudiArabia floods.	قناة تقهبر التركية الفيضانا ت تضرب المملكة العربية السعودية.	tvahaber channel Turkish Floods hit Saudi Arabia.
#SaudiArabia Sends Aid to Citizens Affected by Floods in #Sudan	فرانس برس: ارتفع عدد ضحايا الفيضانات في المملكة العربية السعودية التي تضر ب البلاد خلال الأيام الثلاثة الأخيرة الى ٧	AFP: The number of flood victims in Saudi Arabia, which hit the country during the last three days to 7
#Venezuela could collapse and take much of its #oil production with it or summer can end and #SaudiArabia really floods us. Good Luck!	إشعاعات واجهزه لتفريق السحب من إيران تت سبب فى الفيضانات فى المملكه العربية الس عودية مراجعلهاتسقط رافعة الحرم المكى	Radiation devices to disperse the clouds of Iran cause floods in Saudi Arabia Mrajolhatsagt crane Haram
Saudi Arabia floods global market destroys American jobs. #OPEC	بداية الفيضانات المفاجئة أمس في المملكة العربية السعودية، ومياه الفيضانات تغمر الجسر في ثوان!!	The beginning of the flash floods yesterday in Saudi Arabia, and flood waters submerged the bridge in seconds !!

Social Media Platforms

JORD collects multi-modal information (images, videos and text) from multiple platforms of the social media. In the current implementation, we retrieve disaster-related images from two platforms, namely Flickr and Google. On the other hand, text and videos are collected from Twitter and YouTube, respectively. It is to be noted that queries in local languages, generated by our system during query refinement, are supported by Twitter, YouTube! and Google, while Flickr only supports queries in English. Figure 4 and Figure 3 provide some sample tweets and images retrieved by JORD for the considered events, respectively.

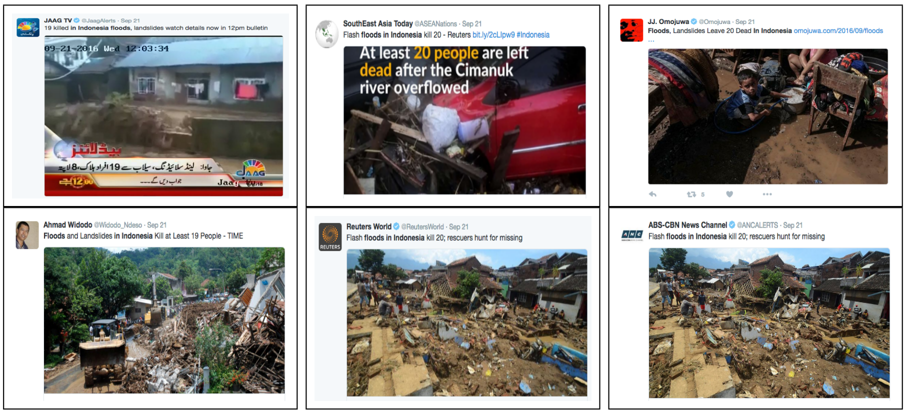


Fig. 4 Sample tweets about recent floods in Indonesia retrieved by JORD.

Fig. 5 A collection of sample images related to natural disasters retrieved from social media through JORD.

Multimedia Content Filtering and Analysis

Among the retrieved content it is important to investigate the relevance of the information collected. In this section, we provide a detailed description of the methodologies we propose for content analysis of the collected multimedia data.

Content analysis of the retrieved images

The basic motivation for content analysis is to filter out irrelevant or less informative images, and limit the results only to the ones, which well represent the underlying events. To this aim, we perform explorative multi-class recognition experiments on the images collected by JORD. We have first created a dataset from the most common disasters (cyclone, drought, earthquake, floods, thunderstorm, tornado and wildfires). As negative samples, we also populate the dataset with another class, namely not-relevant which includes the non-relevant images. The dataset is then divided into training and test sets by choosing the images for both sets randomly.

From the implementation point of view, we use two main approaches, relying on two different types of image representation schemes including: (i) classification using global features (GF), and (ii) classification using features extracted with multiple pre-trained deep models. In the GF approach, we rely on Lire [37], and extract JCD features with a feature vector of size 167. In the second approach, we relay on deep models based on it outstanding performance in other application domains[60, 61], and extract features from each image through 7 different models, pre-trained on ImageNet [18] and places dataset [64], from 4 different deep architectures. These architectures include AlexNet [33], GoogleNet [54], VGGNet [51] and ResNet [26]. For feature extraction with AlexNet, VGGNet, and ResNet we made use of the Caffe toolbox[11]; for GoogleNet we relied on Vlfeat Matcovnet[12]. AlexNet and VGGNet returned a feature vector of size 4096; GoogleNet and ResNet (all configurations) provided feature vectors of size 1024 and 1000, respectively. In this work, we simply use the pre-trained models to extract features from an image without any training or fine-tuning, and the extracted features are then fed as input to the classifiers, which provide results in the form of posterior probabilities. Subsequently, these results are combined in a late fusion method.

Content Analysis of the retrieved tweets

The objective of tweets analysis is two-fold. On one hand, we assess the quality of the retrieved tweets and filter out the less informative and irrelevant tweets. On the other hand, we are interested in collecting the coordinates of the areas affected by the disaster. To do so, we perform the following two experiments on the collected tweets.

- We perform binary and multi-class classification to identify and filter-out the irrelevant tweets. This experiment is intended to improve users' experience providing them with more appropriate data.
- We also analyze and extract the places and city names mentioned in the tweets' text to retrieve and link satellite imagery with the events. The basic motivation for this experiment comes from the fact that the GPS coordinates associated with tweets do not necessarily match with the location of the disaster. Moreover, the presence of GPS information in tweets is not always guaranteed. Instead, users tend to mention the exact places affected by the underlying disaster in the text.

Similarly to image analysis, we started with the collection of a dataset by choosing tweets related to eight common natural disasters from the pool of tweets retrieved by JORD.

[11] http://caffe.berkeleyvision.org/
[12] http://www.vlfeat.org/matconvnet/

These disasters include: cyclone, drought, earthquake, floods, landslides, snow-storm, thunderstorm and wildfires. We also populate the dataset with 8 additional classes including: not-relevant-cyclone, not-relevant-drought, not-relevant-earthquake, not-relevant-floods, not-relevant-landslides, not-relevant-snowstorm, not-relevant-thunderstorm and not-relevant-wildfires. For the labeling of tweets with positive and negative samples (i.e., relevant and irrelevant tweets), annotation is performed manually. To further populate the negative samples, we crawled twitter with additional queries containing the names of the countries affected by the disaster.

To discard irrelevant tweets, we have explored two different solutions: (i) binary classification (i.e., relevant vs non-relevant), and (ii) multi-class classification with 9 classes: 8 of them refer to disaster events, while the 9th represents the non-relevant tweets. As far as the text analysis is concerned, we rely on a state-of-the-art library[13], used both for tweets' classification and to retrieve places and city names. Initially, text is broken into tokens, followed by identifying the places and city names in the extracted tokens with the help of an internal database maintaining the list of places and cities of different countries. Some sample tweets, where the places (e.g., states, districts, city and local areas names) affected by the underlying disaster are mentioned, include : "FIF Pakistan distribute Relief goods of Drought victims in Tharparkar (city name)", "The EU supports livelihoods nutrition in drought-stricken Sindh (province name) Pakistan", "11 dead 50 wounded in Bundibugyo (District name) landslide Uganda". Moreover, the GPS coordinates of the identified places and cities are crawled for remote sensed imagery, which are then processed for disaster detection.

Linking social media with remote sensed data

In this section we detail how the geo-location information is used to retrieve and link remote-sensed images to the underlying events. To this aim, JORD relies on Google Earth, which provides satellite images continuously; this allows to retrieve series of images before and after a disaster. JORD extracts GPS information from the retrieved data (images and tweets) and crawls Google Earth over a time window centered in the event date. Figure 6 shows sample satellite images of the national palace of Haiti retrieved through Google Earth before and after the earthquake. Without loss of generality other sources of remote sensed data can be crawled and integrated in JORD. Figure 7 shows a sample output of our system for a query about recent floods in Kenya, where the retrieved images, tweets, videos and the satellite data from Google Earth are shown.

Flood detection in satellite images: a use-case

In this section, we present the application of our method to the use case of flood detection in satellite images. The proposed method is designed to process image patches of satellite images covering a wide spatial areas of multiple instances of flooding events. The satellite image patches are usually recorded during (or shortly after) the flooding event at different locations. The basic satellite imagery used in this work has been taken from Planet's 4-band satellites [56]. The whole dataset and the corresponding data usage instructions are publicly available [11] and consists of a set of image patches with corresponding pixel-level segmentation masks of the flooded areas. The image patches are stored in the non-normalized 4-channel 16-bit TIFF file format while the corresponding segmentation masks are stored in the 1-channel 8-bit PNG file format.

[13] https://textblob.readthedocs.io/en/dev/

Fig. 6 Sample Google Earth images before and after Haiti earthquake.

The image patches consist of fours 16-bit channels: Red (R), Green (G), Blue (B) and Infrared (IR). None of the existing satellite image visualization software was able to display such data correctly. Moreover, most of the existing image processing software are designed to be used with standard three-channel RGB images. To overcome this issue, we decided to convert each image patch into a pair of images, namely three-channel RGB and single-channel IR images. After the extraction of raw channels data, we performed the normalization for both image components, independently. For the RGB images, we use the joint three-channel normalization, which fits all the R, G and B pixel values of the input geo-image into the standard 0-255 RGB values region. It has to be noted that the normalization coefficients are kept the same for all three channels, which helps to achieve real color balance even in cases of low variations in one of the three components. The normalization of the IR component is performed separately, as shown below:

$$rgb_{min} = \min(\min_{i \in R} r_i, \min_{i \in G} g_i, \min_{i \in B} b_i)$$

$$rgb_{max} = \max(\max_{i \in R} r_i, \max_{i \in G} g_i, \max_{i \in B} b_i)$$

$$ir_{min} = \min_{k \in IR} ir_k, \quad ir_{max} = \max_{k \in IR} ir_k$$

$$\forall i \in \{R|G|B\} \quad \{r|g|b\}^*_i = \frac{255(\{r|g|b\}_i - rgb_{min})}{rgb_{max} - rgb_{min}}$$

$$\forall k \in IR \quad ir^*_i = \frac{255(ir_k - ir_{min})}{ir_{max} - ir_{min}}$$

$$\forall i \in (R \cap G \cap B) \quad gray_i = 0.299 r_i + 0.587 g_i + 0.114 b_i$$

After the conversion to RGB and IR image pairs (see example in figure 8), we have performed a visual analysis of the converted images in order to assess the resulting image quality, the correctness of the conversion and the contents of the dataset. We have found images

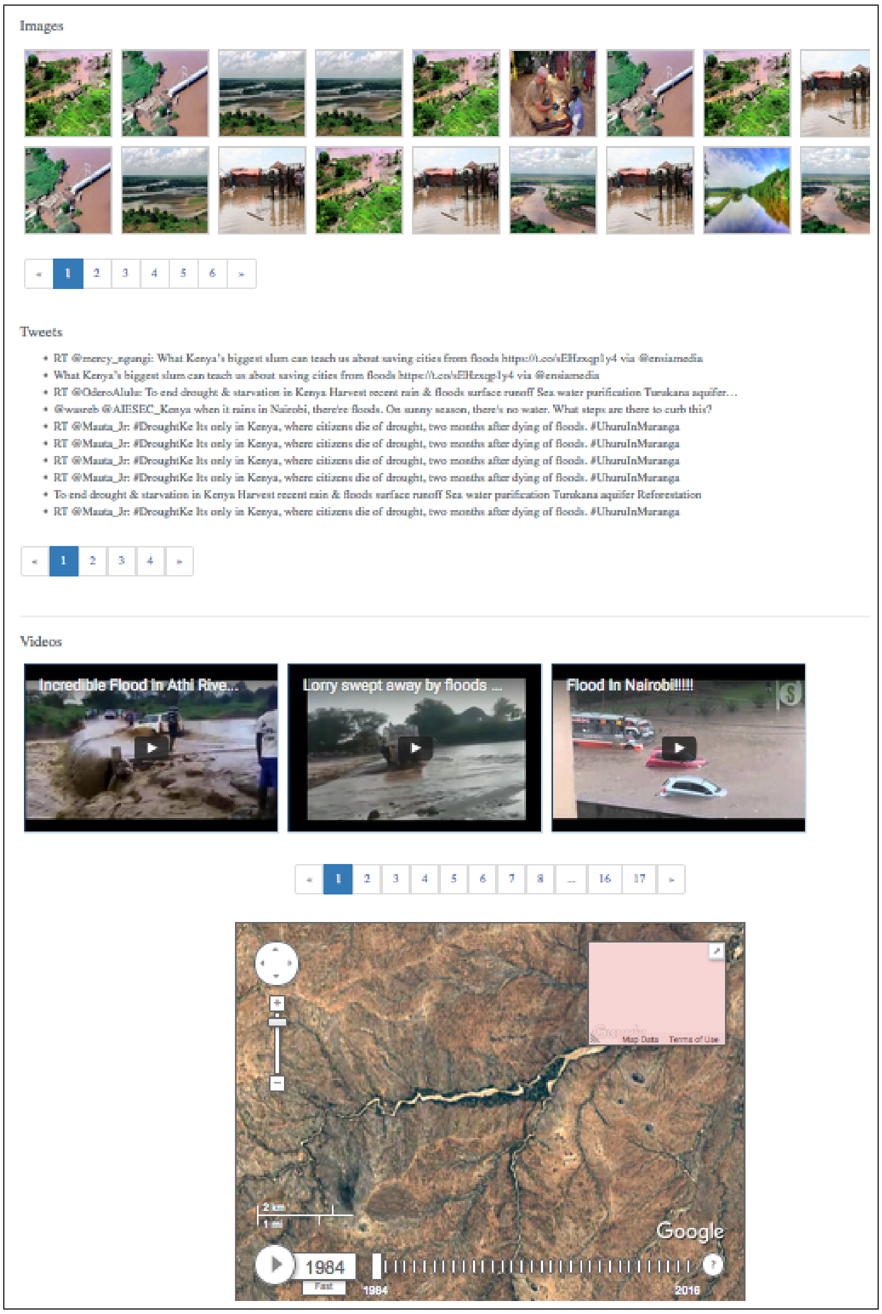


Fig. 7 A sample output of JORD in terms of retrieved images (at the top), tweets, videos and the corresponding satellite data.

to be non-contrast, blurry and significantly color-range-limited. During our initial experiments, we realized that it is not possible to use off-the-shelf image segmentation frameworks due to the nature of the provided satellite imagery. Based on our previous experience [47],

(a) Three-channel normalized RGB image. (b) Single-channel normalized IR image. (c) Flooding area segmentation mask.

Fig. 8 Example of the converted image patch from the original satellite imagery.

we decided to use GANs as the main segmentation method. GANs [23] are machine learning algorithms used in unsupervised learning, and implemented via two neural networks contesting with each other in a zero-sum game framework. They achieved promising results both in terms of performance and data processing speed in image segmentation tasks.

As the basis for our method, we use a neural network architecture originally developed for the retinal vessel segmentation in fundoscopic images with GANs (V-GAN) [52]. The V-GAN architecture [52] is designed for the processing of retinal images that have comparable visual properties, and provides the required output with one-class per-pixel image segmentation output.

In order to adapt V-GAN to our flood detection approach, we modified the network architecture by changing the top-layers configuration in order to support both standard three-channel RGB and four-channel RGB+IR geo-image-compatible input. Furthermore, the final layer of the generator network is extended with a threshold activation layer to generate the binary segmentation maps.

During our initial experiments with our model, we observed that, though the modified V-GAN is able to perform the segmentation of the provided satellite images, the estimated performance metrics were below the expected level. Additional visual analysis of the converted RGB and IR images showed that sometimes the IR component of the sourced geo-images is irrelevant to the flooding areas, which is one of the possible reasons that caused our model to be biased during the training process, preventing it from the extraction of the properties of the flooding areas. Based on these considerations, we decided to exclude the IR component from the model input, and process the RGB components only, which resulted in a good detection performance. Furthermore, we continued to investigate deeper into the multi-channel approach and, after debugging of our model, we realized that the used normalization scheme is causing problems. Despite the good results obtained from the detection using the RGB-normalized images only, the independent normalization procedure of IR channel was resulted in the significantly base-value-shifted output images, mostly because of the high variations in the IR channel caused by the significant difference of the value of the reflected IR light depending on the day time and cloud coverage for the area. To resolve this problem, we redesigned the data preparation and augmentation code as well as the input layer of our model in order to support direct input of the raw satellite imagery data. The best-performing data augmentation scheme implements rotation and flipping of the initially prepared RGB and IR frames stored in floating point format. Preparation is performed by masking out all the over- and under-exposed frame pixels with the numerical values outside of $1\% - 99\%$

range computed for the GRAY and IR component independently. Rotation is performed with $20°$ steps for the original and the flipped frames, resulting in 35 new frames complementary to the original ones. Frames preparation was used for both training and testing, but frames rotation and flipping were used only for model training. This resulted in significant improvements in the model training behaviour and allowed us to perform experimental evaluations using both RGB and RGB+IR channels configurations.

4 System Evaluation via Crowdsourcing

To evaluate the system in terms of correctness of the retrieved multimedia contents and usefulness for the users, we also conducted two crowdsourcing studies with a large number of workers on Microworker[14]. The workers on the platform were asked to give their opinion about the multimedia contents retrieved, and about the system itself. The first study contained 6 questions. After we analyzed the collected data, we decided to run a second study to verify the output of the first one. For the second study, we slightly modified 2 questions for better understanding (but with the same goal), and in addition we, added 5 questions to increase the insight.

In order to assure the quality of the crowdsourcing task, workers were paid .75 USD per activity, and we tried to be as fair as possible regarding the discarding of workers. Considering the fact, as also shown in [49], that controlling and discarding workers can lead to an undesired outcome of the study, we tried to accept almost every worker if they did the task in a proper way. We only discarded workers that clearly did not perform the task (blank questions, very short time to complete it).

Figure 9 depicts the design of the task proposed for the evaluation of the JORD through crowdsourcing. In the proposed task, an event is represented to the workers by providing three different types of multimedia contents including images, tweets and videos. We asked the workers eleven different questions:

(i) Do you think the information provided by the proposed system is useful and detailed enough to cover a detailed story of the underlying disaster? This question aims to get feedback from workers about the usefulness of the collected multimedia data using a scale from one (not useful) to five (very useful).
(ii) What was the main cause of the event? (please answer with a single word for example earthquake, storm, etc.). This question is used to evaluate if JORD can help the user to understand the cause/type of the natural disaster or not.
(iii) From three possible events, which one do you think has been the one presented to you? This question is used to evaluate if JORD can help the user to understand the retrieved event or not.
(iv) How useful was each type of information for you? Here, the worker had to scale the usefulness of different type of multimedia content (images, tweets and videos) from one to five from not useful to very useful.
(v) I found the various sources of information in the proposed system well integrated and useful. In this question, the users are asked to share their experience about the integration of information collected form different sources. (additional question)
(vi) I would imagine that the general public would find the system useful. This question aims to ask the users to rate the usefulness of the system from general public view point. (additional question)

[14] https://ttv.microworkers.com

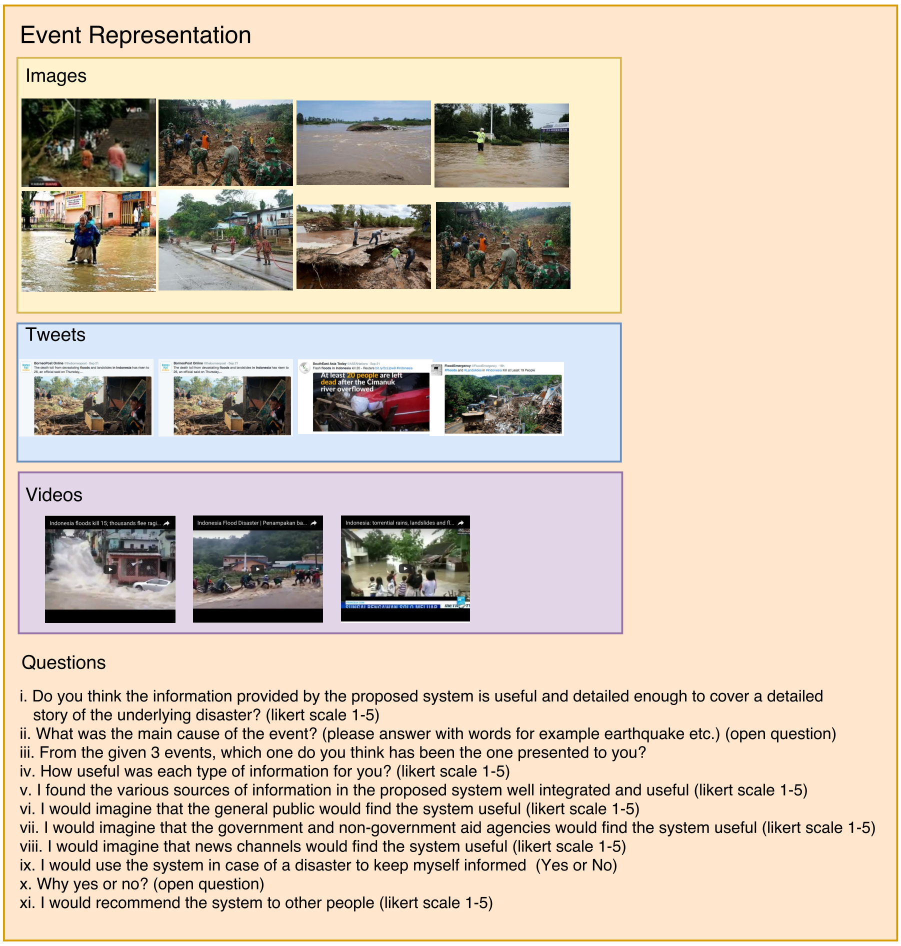


Fig. 9 The design of the crowdsourcing task developed for the evaluation of the JORD system. At the top, an introduction to the task is provided with details of the proposed system. An event is represented with three different types of information including images (from Flickr and Google images), Tweets (from Twitter) and videos from Youtube. Ten different questions are posed regarding the event and importance of the provided information.

(vii) I would imagine that the government and non-government aid agencies would find the system useful. In this question, we ask the users to present their views about the potential of the system to be used by the government and non-government aid agencies. (additional question)

(viii) I would imagine that news channels would find the system useful. In this question, we want to investigate how much the system can be helpful for the news agencies especially in areas where they don not have reporters. (additional question)

(ix) I would use the system in case of a disaster to keep myself informed. A simple question to be answered with yes or no depending on their opinion about the system.

(x) Why would you use or not use it? This was an open question where the workers had to reason their yes or no from the previous question. This question is used to assure the quality of the workers responses. The responses by the workers who did the task in a wrong way are filter out based on the answers in this question. Each answer is investigated manually, and if the answer make sense and showed that the worker did it in a proper way, we accepted it, otherwise removed from the final evaluations.
(xi) I would recommend the system to other people. This question simply aims to investigate the overall recommendation of the users. (additional question)

5 Results and Discussion

In the next subsections, we provide experimental results and a detailed analysis of the experiments conducted on images, tweets, videos and satellite imagery along with the detailed statistics of the crowdsourcing study.

Content based analysis

Content based analysis of retrieved Images

In this work, to differentiate among relevant and non-relevant images retrieved by JORD, we perform two different experimental configurations. In one set of experiments, we rely on individual features extracted through Lire library (Global Features) and different deep models. The underlying insight of this experiment is to analyze and evaluate the performances of different features descriptors in the context of content analysis of disaster related images. In the second experiment, based on our previous experience [5], we combine the classification results of different deep models in a late fusion method. It is to be noted that we do not perform any data augmentation, such as cropping, for any of the approaches mentioned earlier. Experimental validations of both approaches are carried out on the images retrieved with JORD system where separate training and test sets are used.

Table 4 shows the experimental results of our content analysis scheme with individual features. Over all, in terms of accuracy, better results are reported with VGGNet pre-trained on places datasets. During our analysis, we observed better results for the models pre-trained on places datasets compared to the ones pre-trained on ImageNet datasets. It is to noted that the models pre-trained on ImageNet correspond to object-level information while the ones pre-trained on places datasets extract scene-level features from an images. The experimental results reveal the importance of scene-level features over the object-level information in the analysis of natural disaster images. Based on our previous experience [5], we believe that object and scene-level information well complement each others in such applications. Therefore, in our next experiment, we combine these models in a late fusion scheme to obtain better classification results.

Table 6 shows the experimental results of our second experiment, where we analyze the performance of our content analysis scheme by coming different deep models relying on late fusion. All in all, we used three different combinations including (i) combination of all 7 models; (ii) models pre-trained on places datasets (VGGNet and AlexNet), and (iii) top two performing models (VGGNet pre-trained on Places dataset and ResNet-152). In the current implementation, we use equal weights for all models in the fusion. Over all, better results are reported by combining all the models achieving an overall accuracy of 78%,

Table 4 Classification results of our content based analysis of images retrieved by JORD with individual features

Events	**Accuracy with different Features**							
	Global Features (JCD)	AlexNet (ImageNet)	AlexNet (Places)	VggNet (ImageNet)	VggNet (Places)	GoogleNet	ResNet-50	ResNet-152
Cyclone	.041	.239	.350	.273	.367	.188	.247	.256
Drought	.209	.586	.641	.606	.620	.544	.758	.751
Earthquake	.640	.748	.798	.809	.832	.721	..827	.818
Floods	.631	.651	.740	.696	.789	.685	.675	.634
Landslides	.253	.444	.398	.481	.490	.425	.592	.574
Snow-storm	.132	.611	.820	.746	.791	.611	.850	.850
Thunder-storm	.753	.848	.864	.867	.848	.846	.878	.878
Wildfires	.468	.650	.674	.773	.716	.596	.711	.758
Non-relevant	.363	.648	.726	.657	.681	.636	.665	.689
Overall	**.513**	**.668**	**.720**	**.726**	**.739**	**.651**	**.728**	**.731**

which shows that these models complement each others. This fact can also be verified by our participation in a benchmarking challenge on MediaEval 2017, where the classification score from different CNN models is combined in a late fusion scheme, achieving first place on run 1 with visual features only [5,3]. We observed better results with certain disasters, such as earthquakes, wildfires and floods, which is mostly due to the fact that these events have specific textures and patterns, and thus can be better identified and recognized through visual content. However, the images related to cyclone, where the accuracy is around 23%, less likely to posses any specific texture, and thus are sometimes very difficult to recognize or differentiate solely through visual content. Nevertheless, we observed that using the visual content for filtering the retrieved images can be a promising step as also depicted by our results.

We also provide comparison of the method against state-of-the-art methods proposed for the MediaEval challenge on Disaster Images Retrieval from Social Media (DIRSM) task. The challenge covered flood related images, only. In Table 5, we provide comparisons of the methods proposed for the the task on run 1, which is purely based on visual information, and average precision at different cutoffs is used as an evaluation criterion. We also participated in the challenge with two different teams, namely UTAOS and MLDCSE, and obtained first and second places on run 1, respectively. In this work, we use different implementation of ResNet models compared to the ones we used in our work with team UTAOS, as detailed in Section 3. The proposed approach achieves an overall gain of 5.65% at cutoff 480 and 2.09% at different cutoffs (50,100,150,240 and 480) over the highest score (our team [5]) in the challenge using visual features only.

Table 5 Comparisons against state-of-the-art on the MediaEval benchmark dataset in terms of average precision at different cutoffs (50,100,150,240 and 480), and precision at 480

Method	Precision		Method	Precision at different cut-oofs	
	Avg. Pre. at diff. cutoffs	Pre. at 480		Avg. Pre. at diff. cutoffs	Pre. at 480
WISC [43]	.6275	.5095	Zhengyu et al. [63]	.6470	.5146
Lopez et al. [34]	.7016	.6158	Hanif et al. [42]	.8098	.6500
ELEDIA [41]	.8787	.7762	UTAOS (our) [?]	.9511	.8494
Konstantinos et al. [32]	.9227	.7882	MLDSCE [5]	.9573	.8681
Keiller et al. [31]	.8788	.7460	Proposed Method	.9782	.9246

Table 6 Classification results of our content based analysis of images retrieved by JORD with different combinations Deep Models

Events	**Accuracy and Precision with different combinations of Deep Models**					
	All Models		Both Places		Top 2 (VggNet places+ResNet152)	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
Cyclone	.239	.510	.299	.448	.324	.443
Drought	.744	.613	.641	.574	.731	.585
Earthquake	.863	.841	.843	.789	.866	.832
Floods	.755	.751	.774	.728	.749	.735
Landslides	.574	.574	.490	.540	.555	.566
Snow-storm	.850	.904	.880	.893	.865	.935
Thunder-storm	.931	.721	.892	.701	.892	.849
Wildfires	.773	.936	.768	.913	.789	.913
Non-relevant	.775	.887	.742	.883	.726	.890
Overall	**.780**	-	**.764**	-	**.774**	-

Table 7 binary tweet-classification results with Naive Bayes and Decision Tree classifiers

Disasters	**Naive Bayes**				**Decision Tree**			
	Acc.	**Prec.**	**Recall**	**F1**	**Acc.**	**Prec.**	**Recall**	**F1**
Cyclone	.9322	.9189	1.0	.9557	.9661	.9705	.9705	.9705
Drought	.6451	.6956	.8421	.7640	.774	.973	.77	.859
Earthquake	.8521	.7972	.9833	.7618	.9565	.933	1.0	.965
Floods	.8971	.851	.9273	.8876	.92	.8659	.9491	.9056
Landslides	.9271	.9176	.975	.9457	.8344	.851	.862	.8569
Snowstorm	.8785	.8387	1.0	.9126	.8598	.8431	.347	.886
Thunderstorm	.8095	.7441	.8648	.8003	.8095	.729	.843	.7818
Wildfires	.6744	.75	.78	.7352	.9065	.92	.92	.92

Content based analysis of retrieved Tweets

As described above, to filter out the irrelevant tweets, we perform two different experiments, namely (i) binary classification (relevant vs. non-relevant for each event, independently) and (ii) multi-class classification (considering 8 events and one class with non-relevant tweets). In both experiments, we rely on two different classifiers namely, NavieBayes and Decision Tree classifier provided in TextBlob toolbox[15].

Table 7 provides experimental results of our binary classification experiment, in terms of accuracy, precision, recall and F-measure, with NaiveBayes and Decision-Tree classifiers. Over all, better results are obtained on each disaster with both classifiers, which shows that content analysis of tweets helps to filter out irrelevant tweets. However, clear advantages of Decision Tree classifier over NaiveBayes can be observed. Moreover, looking at the accuracy, better results are observed for most of the disasters except droughts. One of the possible reasons is in the training dataset and the event name itself. Some sample tweets which are miss-classified for drought include: "A Steve Smith century has left Australia well-placed to end its 13-year test drought in India.", "Fast Fingers crossed: Zimbabwe expected to break cricket drought in Pakistan", and "Will Australia break its drought in India and get the victory? Or will India continue its", etc.,.

Table 8 shows the results of the multi-class classification experiment on the retrieved tweets. As can be observed, better results are obtained on each individual disaster as well as

[15] https://textblob.readthedocs.io/en/dev/

Table 8 Multi-class classification with Decision-tree classifier

Event	Accuracy
Cyclone	.9491
Drought	.7903
Earthquake	.9565
Floods	.901
Landslides	.8741
Snowstorm	.8037
Thunderstorm	.75
Wildfires	.9069
Non-relevant	.912

non-relevant class. It is to be noted that in this experiment the non-relevant class is composed of the individual non-relevant classes of each disaster.

Flood Detection in Satellite Images

For the detailed experimental evaluation of our algorithm of flood detection in satellite images, we used the publicly available dataset [11] of the Multimedia Satellite Task, which was a part of the 2017 MediaEval Benchmarking Initiative for Multimedia Evaluation [16]. The dataset consists of development and validation sets of satellite image patches with corresponding pixel-level segmentation masks of the flooded areas. The development set consists of 463 image patches with corresponding flooding segmentation masks. The test set contains 260 image patches along with the flooding segmentation masks. The dataset covers seven different flooding events occurred in the different regions of the world. In order to evaluate the generalization properties of our detection algorithm, we mixed all the images from the different events and regions interpreting the image sets as the data sources with unknown geographical, temporal and event-related information. In this experimental study, we evaluate our method with a two-fold cross-validation strategy. During a first evaluation run, we used the dataset in the original order: the development set is used as a training set, and the validation set as a test set. In the second evaluation run, we used the dataset in the flipped order: the development set is used as a test set, and the validation set is used as a training set. The non-equal splitting of the number of images in the training and test sets can be seen as an additional test for the redundancy and the efficiency of the proposed detection algorithm. Moreover, we also performed the evaluation of both detection approaches: three-channel normalized RGB and four-channel raw RGB+IR, which gave us four different evaluation runs in total.

The proposed neural network model performs flooded areas detection on the pixel-level, and provides the output in the form of a binary segmentation map, which contains true values (white pixels) for the pixels belongs to the detected flooded areas and false values (black pixels) for the areas without flooding detected. The examples of the model's segmentation output together with the source RGB and IR images as well as corresponding ground truth masks are presented in figures 10, 11 and 12. As one can see, the RGB and IR channels provide a different information about the region being analyzed. In most of the cases, the four-channel combination of RGB and IR channels results in the better detection performance and can increase the detection accuracy significantly (see figure 10 for an example). Nevertheless, in some cases when the IR channel contains data that confuses the detection

[16] `http://www.multimediaeval.org/mediaeval2017/`

(a) The image patch. RGB channels (b) The image patch. IR channel (c) Ground truth mask. (d) Segmentation results for three-channel RGB. (e) Segmentation results for four-channel RGB+IR.

Fig. 10 Example of the correctly found flooded area. This example shows that detection was performed better for combination of RGB and IR channels.

(a) The image patch. RGB channels (b) The image patch. IR channel (c) Ground truth mask. (d) Segmentation results for three-channel RGB. (e) Segmentation results for four-channel RGB+IR.

Fig. 11 Example of the correctly found flooded area. This example shows that in some cases detection was performed better for three RGB channels.

(a) The image patch. RGB channels (b) The image patch. IR channel (c) Ground truth mask. (d) Segmentation results for three-channel RGB. (e) Segmentation results for four-channel RGB+IR.

Fig. 12 Example of the false positive detection of the flooding. The water in this image patch is "legal" water in the irrigation channels. To be able to deal with such cases the comparative time-based analysis must be added to the detection algorithm utilizing many satellite images of the same region taken in different periods of time.

algorithm and leads to a mis-detection with a tendency to increase number of false-positive pixels (see figure 11 for an example). Moreover, in some quite rare cases (at least within the dataset used) no combination of the channels are sufficient to perform a distinctive and accurate detection of the flooded areas because of a presence of a water that is "legal" (non-flooded water e.g., lakes and rivers) (see figure 12), for example in the irrigation channels, normal rivers and lakes, etc. To be able to deal with such cases the comparative time-based analysis must be added to the detection algorithm utilizing as many satellite images of the same region taken in different periods as possible. The time-based analysis will be a subject of a future work of our research.

Our proposed model for the flood detection includes the top layer with an adjustable threshold parameter, which is used for the final output segmentation map binarization. The value of this threshold parameter defines a border line for each pixel to be counted as belonging to flooding area depending of the model's output probability value, and it has a direct effect on the number of flooded pixels detected. Thus, in order to perform a complete model evaluation, we have repeated all four evaluation runs with different values of the threshold parameter. For an overall performance evaluation of this threshold-value-effect evaluation experiments, we selected the Matthews correlation coefficient (MCC) which is used in machine learning as a measure of the quality of binary (two-class) classifications. It takes into account true and false positives and negatives, and is generally regarded as a balanced measure that can be used even if the classes are of very different sizes. The MCC value lies in the region between -1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and 1 indicates total disagreement between prediction and observation. Our previous research [46] confirmed that MCC is the most convenient metric for the binary classification tasks.

The results of the threshold value evaluation are depicted in figure 13. As one can see, the low < 0.1 and the high > 0.9 values of the threshold have strong negative effect on the performance of the proposed model. The optimal threshold value lies, as it was expected before the experimental studies, in between 0.4 and 0.6 depending on the exact order of the samples in the dataset and the number of channels used. The best value of the threshold parameter is 0.42 for the three-channel RGB model regardless of the sets order. For the four-channel RGB+IR model, best values are 0.518 in the case of the original test and train datasets, and 0.53 for the flipped order. Despite the fact that the best values of the threshold parameter are slightly different depending on the dataset and number of channels used, the resulting difference of the performance is small for the threshold values within the interval from 0.4 to 0.6 (which includes the found best values) and for the future work we will use the threshold value of 0.5 for all the cases. Nevertheless, in this work we have performed an evaluation of the performance metrics of the developed detection method using the best found threshold values. The interesting finding is that RGB and RGB+IR approaches perform almost equally for the original dataset, but RGB+IR performs better for the flipped datasets. That can be caused by a significantly reduced size of the training dataset in the flipped runs, which makes one additional information channel important for the proper model generalization during training process.

The results of the performance evaluation for the best threshold values are presented in table 9. The first two runs was performed by the three-channel RGB model and the original and flipped datasets. The threshold value used was 0.42 for both runs. The results show that the best MCC evaluation performance metrics value of 0.805 was achieved for the original datasets order. For the flipped datasets order, MCC metrics was a slightly lower with a value of 0.742 which can be caused by the significantly lower number of the training images in the flipped datasets that caused a less level of the model generalization. Nevertheless, the MCC values as well as other common performance characteristics depicted in table 9 confirms the validity and usability of the model developed together with the high adaptation rate and ability to learn even on the limited training dataset size.

The performance results computed for the four-channel RGB+IR model runs (see table 9) with the threshold value of the 0.518 and 0.53 for the original and flipped datasets respectively shows almost the MCC performance values of 0.8 and 0.758 respectively. As one can see, the original datasets run has the same performance as the RGB model (the difference is not significant). For the flipped run the RGB+IR model perform noticeable better, thus using of four-channel RGB+IR model can be considered as preferable for the flooding detection tasks. Moreover, visual inspection of the datasets provided showed that in some cases IR channel may provide distinctive clues for distinguishing between flooded areas and

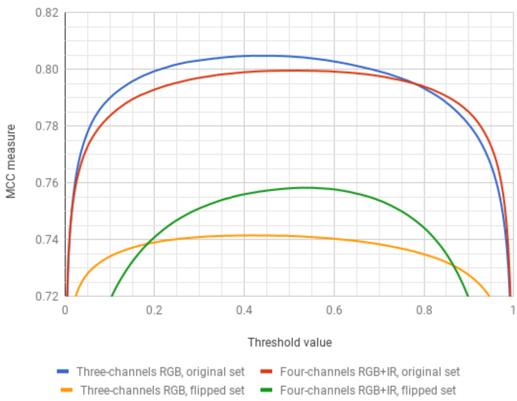


Fig. 13 The comparison of the flooding detection performance in terms of MCC measure computed with the different probability threshold values for three- and four-channel satellite imagery data for the original and the flipped datasets.

"normal" water areas, but this should be investigated deeper using more datasets of bigger sizes.

Table 9 Two-fold cross-validation results for the two presented flooding detection approaches. The performance numbers of accuracy (ACC), precision (PREC), sensitivity or recall (REC), specificity (SPEC), F-Measure (F1) and Matthews correlation coefficient (MCC) are presented in the original / flipped order regarding to the original dataset [11] for the selected values of the probability threshold value (THRESH).

Input data	THRESH	ACC	PREC	REC	SPEC	F1	MCC
Three-channel normalized RGB	0.420 / 0.420	0.913 / 0.883	0.879 / 0.835	0.861 / 0.827	0.940 / 0.913	0.870 / 0.831	**0.805 / 0.742**
Four-channel raw RGB+IR	0.518 / 0.530	0.911 / 0.889	0.883 / 0.827	0.849 / 0.862	0.943 / 0.904	0.865 / 0.844	**0.800 / 0.758**

Crowdsourcing Analysis

The preliminary statistics of the first study have been reported in our earlier work [4]. Here, we provide more detailed analysis of the conducted crowdsourcing study and also discuss the second study.

Crowdsourcing Study I

For the first study, 385 responses from the workers were collected. Based on analysis of the open questions, we discarded 36 responses showing that either they did not understand the task properly or did it in a wrong way (tried to cheat or did not take it serious). The statistics of the remaining 349 valid responses are provided in Table 10 in terms of valid responses per event. In total, 51 events were presented to the workers, and for each event, we got at least 5 valid responses.

The first question (i), where the workers had to state if they find the system useful or not, had an average of 4.47 for all workers. This is a clear indication that workers find the system and the provided information useful.

For the second question (ii), where the workers had to chose the correct event out of three provided options, only 19 workers out of 349 failed to correctly recognize the event presented to them. A closer investigation showed that all of them had just the country wrong but gave a correct answer about the disaster. This shows two important things. Firstly, that the retrieved information of JORD is accurate and can help users to get more information about events, and secondly, that connecting it to satellite images is important to improve understanding of the event in terms of location.

For the third question (iii), where the workers had to report how useful they found different types of multimedia content, we got an average of 4.23 for images, 4.08 for tweets and 4.44 for videos. Based on this, we can see a tendency that users find videos most useful and tweets least. We think that this might be due to the fact that a video usually contains more information than a text or image and that it helps people more to understand and experience the current situation.

The last two questions (iv) and (v) are evaluated together. For the first question (iv), 336 from 349 workers (around 96%) stated that they find the system useful. This is a promising indicator that such a system would be useful and used by users.

Having a closer look into the answers of the last question (v) from crowdworkers that did not find it useful revealed the following reasons: they find the system scary; can use Google; would use a system that can predict events; or not do and never will face such an event. Examples from users that would like to use the system are: "Yes I would definitely. It would be very useful for people in the affected areas.", "i really want to know how much destructive the natural disaster was and much more related on it.", "It is an interesting way to view news; Videos are always more impressive than images and tweets; I would like to use it to get better trusted info; and It is informative and gives a very COMPLETE view of what is happening". "I love the use of ALL forms of information, as in, photos, videos, and tweets".

Based on our evaluation using crowdsourcing, it appears that such a system would be interesting and useful for users.

Crowdsourcing Study II

For the second study, 771 responses from the workers were collected on 53 different events with at least five different workers per event. Based on analysis of the open questions, we discarded 8 responses showing that either they did not understand the task properly or did it in a wrong way (tried to cheat or did not take it serious).

The results of the 6 questions from our first study are comparable. For example, the average usefulness rate of the overall system in the new study is 4.39 out of 5 in question (i). Similarly, in question (ii), 93.5% of the workers were able to correctly recognize the event from the provided information, and the videos are rated more useful compared to the images and tweets in question (iv). Moreover, 94% of the workers liked to use the proposed system to get information about natural disasters in question (ix). This basically confirms our conclusions drawn based on the first study.

For the first additional question (v), where the workers had to rate how efficiently the information from different sources are integrated in the system, where an average score of 4.32 out of 5 is obtained. Similarly, for the questions (vi), (vii) and (viii), where the workers had to rate the usefulness of the JORD system from the point of view of a general public, aid

Table 10 Statistics of the crowdsourcing study I. The final dataset, after discarding cheating workers, we have in total of 349 distinct responses from 349 different workers.

Event	# Responses	Event	# Responses
Cyclone Roanu Bangladesh	10	Floods in Somalia	5
Cyclone Winston Fiji	7	Floods in Southern China	9
Drought in India	10	Landslides in Myanmar	5
Wildfires in Greece	10	Landslides in Srilanka	5
Wildfires in spain	10	Landslides in Uganda	7
Floods in Ibaraki Japan	6	Landslides in Rwanda	8
Floods in Indonesia	5	Snowstrom in USA	8
Thunder strom in USA	6	Tornadoes in Oklahoma USA	6
Wildfires in California	8	Earthquake Ecuador	5
Floods in Nigeria	6	Thunderstrom in Bangladesh	6
Plane Crash in Sudan	7	Drought in Pakistan	6
Plane Crash in Papua New Guinea	7	Earthquake in Pakistan	7
Earthquake in China	6	Explosion in a plant in Mexico	5
Floods in Kashmir	6	Floods in Port-au-Prince Haiti	7
Earthquake in Japan	5	Earthquake in Spain and Moroco	7
Floods in Chile	5	Floods in Saudia Arabia	5
Floods in Srilanka	10	Drought in Timor Leste	5
Floods in KpK Pakistan	5	Thunderstorm in Myanmar	6
Wildfires in Alberta Canada	5	Earthqauake in Tainan Taiwan	6
Landslide in Sibolangit Indonesia	10	Tornado in Uruguay	11
Storm Katie in France and UK	6	Landslide in Nepal	6
Mudslide in Taining China	5	Coal mine explosion in Ukrine	7
Severe Weather in Haiti	6	Earthquake in Pakistan and Afghanistan	5
Floods in Texas USA	11	Floods in Assam India	6
Heat Wave in India	6	Floods in Ethiopia	6
Cyclone Champala Yemen	6	**Total**	**349**

agencies and news channels. The obtained average scores for these questions are 4.41, 4.28 and 4.38, respectively. Based on these scores, we can see that the proposed system is rated satisfactory useful for different users. For the final question (xi), where the workers had to give an overall recommendation for the proposed system, an average score of 4.57 has been obtained, which is a promising indication that the system is highly recommended and would be used by people if made openly available.

Some examples of feedback from workers who find the system useful and would like to use it are: "Yes I would use it because the system of this type provides wider picture of the event."; "I would like to use it to be aware of whats happening and be able to help if I can."; "it is an effective way to spread information by including various forms of media"; " i will use it because the system is very useful, you can find information from various sources at one place".

All in all, more than 1,000 unique people participated in the conducted crowdsourcing studies to evaluate the usefulness of the proposed system and the information it provides. Based on these two studies, we can conclude that our JORD system would be well received by people and also used. One of the main positive feedback we could observe was the fact that people got different sources and medias of information presented in a structured and easy way. Especially, in context of emergencies people seem to like a good overview and diverse information representation.

6 Conclusions and Future Work

In this paper, we have presented our system, JORD, to autonomously and automatically retrieve multimedia information from social networks about natural disasters, and link it to satellite imagery. Moreover, we performed content analysis of the retrieved multimedia data to provide more relevant information to users in the form of text, images and videos. We showed that the combination of social media data and satellite images provide a more detailed story of the underlying disaster events. We also demonstrated that queries in local languages that are relevant to the exact position of natural disasters retrieve more accurate information about a disaster event.

The evaluation of the JORD system was also carried out through a crowdsourcing-study, where workers were asked to evaluate the usefulness of the system and to identify an event presented to them with collected images, tweets and videos. The evaluation indicates that JORD works very accurate without human input, and it can be used to collect and merge a large number of event based data for technological and environmental disasters from different sources.

In the future, we plan to include more different social media platforms and also to improve the representation of the retrieved and linked information to the users. We also plan to extend the content analysis to videos retrieved by the JORD system, which will further increase a user's experience and satisfication.

References

1. Ahmad, K., Conci, N., De Natale, F.: A saliency-based approach to event recognition. Signal Processing: Image Communication **60**, 42–51 (2018)
2. Ahmad, K., De Natale, F., Boato, G., Rosani, A.: A hierarchical approach to event discovery from single images using mil framework. In: Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on, pp. 1223–1227. IEEE (2016)
3. Ahmad, K., Konstantin, P., Riegler, M., Conci, N., Holversen, P.: Cnn and gan based satellite and social media data fusion for disaster detection
4. Ahmad, K., Riegler, M., Pogorelov, K., Conci, N., Halvorsen, P., De Natale, F.: Jord: A system for collecting information and monitoring natural disasters by linking social media with satellite imagery. In: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, p. 12. ACM (2017)
5. Ahmad, S., Ahmad, K., Ahmad, N., Conci, N.: Convolutional neural networks for disaster images retrieval. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
6. Amit, S.N.K.B., Shiraishi, S., Inoshita, T., Aoki, Y.: Analysis of satellite images for disaster detection. In: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, pp. 5189–5192. IEEE (2016)
7. Atefeh, F., Khreich, W.: A survey of techniques for event detection in twitter. Computational Intelligence **31**(1), 132–164 (2015)
8. Bakillah, M., Li, R.Y., Liang, S.H.: Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon haiyan. International Journal of Geographical Information Science **29**(2), 258–279 (2015)
9. Bischke, B., Bhardwaj, P., Gautam, A., Helber, P., Borth, D., Dengel, A.: Detection of flooding events in social multimedia and satellite imagery using deep neural networks (2017)
10. Bischke, B., Borth, D., Schulze, C., Dengel, A.: Contextual enrichment of remote-sensed events with social media streams. In: Proceedings of the 2016 ACM on Multimedia Conference, pp. 1077–1081. ACM (2016)
11. Bischke, B., Helber, P., Schulze, C., Venkat, S., Dengel, A., Borth, D.: The multimedia satellite task at mediaeval 2017: Emergence response for flooding events. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
12. Campbell, J.B., Wynne, R.H.: Introduction to remote sensing. Guilford Press (2011)
13. Chang, S.F.: New frontiers of large scale multimedia information retrieval. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, pp. 5–5. ACM (2016)
14. Chavez, P., Sides, S.C., Anderson, J.A.: Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. Photogrammetric Engineering and remote sensing **57**(3), 295–303 (1991)
15. Chen, T., Borth, D., Darrell, T., Chang, S.F.: Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. arXiv preprint arXiv:1410.8586 (2014)

16. Cheong, M., Lee, V.: Twittering for earth: A study on the impact of microblogging activism on earth hour 2009 in australia. In: Springer ACIIDS, pp. 114–123. Springer (2010)
17. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: # earthquake: Twitter as a distributed sensor system. Transactions in GIS **17**(1), 124–147 (2013)
18. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255. IEEE (2009)
19. Du, R., Yu, Z., Mei, T., Wang, Z., Wang, Z., Guo, B.: Predicting activity attendance in event-based social networks: Content, context and social influence. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, pp. 425–434. ACM (2014)
20. Earle, P.S., Bowden, D.C., Guy, M.: Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics **54**(6) (2012)
21. Erol, B., Hull, J.J., Lee, D.S.: Linking multimedia presentations with their symbolic source documents: algorithm and applications. In: Proc. of ACM MM, pp. 498–507 (2003)
22. Fisher, A., Flood, N., Danaher, T.: Comparing landsat water index methods for automated water classification in eastern australia. IJRSE **175**, 167–182 (2016)
23. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680 (2014)
24. Guha-Sapir, D., Below, R., Hoyois, P.: Em-dat: International disaster database. Catholic University of Louvain: Brussels, Belgium (2015)
25. Guille, A., Favre, C.: Event detection, tracking, and visualization in twitter: a mention-anomaly-based approach. Social Network Analysis and Mining **5**(1), 1–18 (2015)
26. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
27. Joyce, K.E., Belliss, S.E., Samsonov, S.V., McNeill, S.J., Glassey, P.J.: A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. Progress in Physical Geography (2009)
28. Jung, M., Henkel, K., Herold, M., Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling. IJRSE **101**(4), 534–553 (2006)
29. Kamilaris, A., Prenafeta-Boldú, F.X.: Disaster monitoring using unmanned aerial vehicles and deep learning
30. Kansas, J., Vargas, J., Skatter, H.G., Balicki, B., McCullum, K.: Using landsat imagery to backcast fire and post-fire residuals in the boreal shield of saskatchewan: implications for woodland caribou management. IJWF **25**(5), 597–607 (2016)
31. Keiller, N., Samuel, F., Ícaro, D., Rafael, W., Javier, M., Otávio, P., Rodrigo, C.: Data-driven flood detection using neural networks. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
32. Konstantinos, A., Anastasia, M., Andreadis, S., Emmanouil, M., Ilias, G., Stefanos, V., Ioannis, K.: Visual and textual analysis of social media and satellite images for flood detection @ multimedia satellite task mediaeval 2017. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
33. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
34. Laura, L.F., Joost, W., Marc, B., Harald, S.: Multi-modal deep learning approach for flood detection. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
35. Li, C., Sun, A., Datta, A.: Twevent: segment-based event detection from tweets. In: Proc. of ACM IKM, pp. 155–164. ACM (2012)
36. Liu, Y., Wu, L.: Geological disaster recognition on optical remote sensing images using deep learning. Procedia Computer Science **91**, 566–575 (2016)
37. Lux, M., Riegler, M., Halvorsen, P., Pogorelov, K., Anagnostopoulos, N.: Lire: open source visual information retrieval. In: Proc. of ACM MMSys (2016)
38. Manjunath, T., Hegadi, R.S., Ravikumar, G.: A survey on multimedia data mining and its relevance today. IJCSNS **10**(11), 165–170 (2010)
39. Mathioudakis, M., Koudas, N.: Twittermonitor: trend detection over the twitter stream. In: Proc. of ACM SIGMOD, pp. 1155–1158 (2010)
40. Meladianos, P., Nikolentzos, G., Rousseau, F., Stavrakas, Y., Vazirgiannis, M.: Degeneracy-based real-time sub-event detection in twitter stream. In: Ninth International AAAI Conference on Web and Social Media, pp. 248–257 (2015)
41. Minh-Son, D., Quang-Nhat-Minh, P., Duc-Tien, D.N.: A domain-based late-fusion for disaster image retrieval from social media. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland

42. Muhammad, H., Muhammad, A., Mahrukh, K., Mohammad, R.: Flood detection using social media data and spectral regression based kernel discriminant analysis. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
43. Nataliya, T., Arkaitz, Z., Procter, R.: Wisc at mediaeval 2017: Multimedia satellite task. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
44. Nurwidyantoro, A., Winarko, E.: Event detection in social media: A survey. In: ICT for Smart Society (ICISS), 2013 International Conference on, pp. 1–5. IEEE (2013)
45. Paul, F., Andreassen, L.M.: A new glacier inventory for the svartisen region, norway, from landsat etm+ data: challenges and change assessment. Journal of Glaciology **55**(192), 607–618 (2009)
46. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference, MMSys'17, pp. 164–169. ACM, New York, NY, USA (2017). DOI 10.1145/3083187.3083212. URL `http://doi.acm.org/10.1145/3083187.3083212`
47. Pogorelov, K., Riegler, M., Eskeland, S.L., de Lange, T., Johansen, D., Griwodz, C., Schmidt, P.T., Halvorsen, P.: Efficient disease detection in gastrointestinal videos–global features versus neural networks. Multimedia Tools and Applications pp. 1–33 (2017)
48. Popescu, A.M., Pennacchiotti, M.: Detecting controversial events from twitter. In: Proc. of ACM IKM, pp. 1873–1876. ACM (2010)
49. Riegler, M., Gaddam, V.R., Larson, M., Eg, R., Halvorsen, P., Griwodz, C.: Crowdsourcing as self-fulfilling prophecy: Influence of discarding workers in subjective assessment tasks. In: 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 1–6 (2016). DOI 10.1109/CBMI.2016.7500256
50. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web, pp. 851–860. ACM (2010)
51. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
52. Son, J., Park, S.J., Jung, K.H.: Retinal vessel segmentation in fundoscopic images with generative adversarial networks. arXiv preprint arXiv:1706.09318 (2017)
53. Stelter, B., Cohen, N.: Citizen journalists provided glimpses of mumbai attacks. The New York Times **30** (2008)
54. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)
55. Takahashi, B., Tandoc, E.C., Carmichael, C.: Communicating on twitter during a disaster: An analysis of tweets during typhoon haiyan in the philippines. Computers in Human Behavior **50**, 392–398 (2015)
56. Team, P.: Planet application program interface: In space for life on earth. san francisco, ca (2016)
57. Tzelepis, C., Ma, Z., Mezaris, V., Ionescu, B., Kompatsiaris, I., Boato, G., Sebe, N., Yan, S.: Event-based media processing and analysis: A survey of the literature. Image and Vision Computing (2016)
58. Wood, H.: The use of earth observing satellites for hazard support: Assessments and scenarios. Final Report of the CEOS Disaster Management Support Group. Available from< http://disaster. ceos. org/legal. cfm (2002)
59. Xu, Z., Zhang, H., Sugumaran, V., Choo, K.K.R., Mei, L., Zhu, Y.: Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media. EURASIP Journal on Wireless Communications and Networking **2016**(1), 44 (2016)
60. Yan, C., Xie, H., Liu, S., Yin, J., Zhang, Y., Dai, Q.: Effective uyghur language text detection in complex background images for traffic prompt identification. IEEE transactions on intelligent transportation systems (2017)
61. Yan, C., Xie, H., Yang, D., Yin, J., Zhang, Y., Dai, Q.: Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE transactions on intelligent transportation systems (2017)
62. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. IJIS **27**(6), 52–59 (2012)
63. Zhengyu, Z., Larson, M.: Retrieving social flooding images based on multimodal information. In: Proceedings of the MediaEval 2017 Workshop. Dublin, Ireland
64. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Proceedings of the NIPS, pp. 487–495 (2014)